Parsimonious Bayesian Factor Analysis with application to milk spectroscopy data

41st Conference on Applied Statistics in Ireland

Alessandro Casa

Joint work with: Tom O'Callaghan & Brendan Murphy

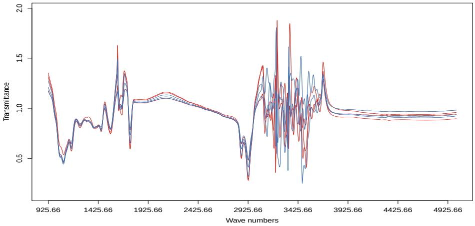
School of Mathematics and Statistics
University College Dublin

alessandro.casa@ucd.ie

13th May 2021

Cows, Diet & Spectroscopy

- Increasing consumers awareness is changing the food industry
- Cattle feeding regimen (pasture vs total mixed ration)
  - pasture diet regarded as more respectful and products as more natural and healtier
  - characterization of the differences implied by different diets on the milk features still overlooked
- Vibrational Spectroscopy techniques
  $\longrightarrow$ cheap, rapid and non-disruptive way to collect vast amount of data + widely used in food science

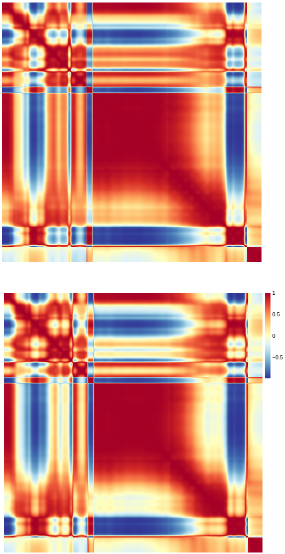


All that glitters is not gold

- Some challenges:
  - High-dimensionality
  - Peculiar correlation structures
- Aim
  - Parsimonious representation
  - Proper wavelengths relationships reconstruction
  - Characterization of the phenomenon

    ↓

    Long story short: extract useful knowledge and insights



Latent variable models - Factor Analysis

- We focus on Factor Analysis (FA)
  - dimensionality reduction...
  - with a focus on the covariance structure
- Let $X = \{x_1, \ldots, x_n\}$ the observed data, FA models $x_i \in \mathbb{R}^p$ as

$$x_i = \Lambda u_i + \varepsilon_i, \qquad i = 1, \ldots, n$$

$\Lambda \in \mathbb{R}^{p \times K}$ loadings, $u_i \in \mathbb{R}^K$ the scores, K the number of factors

$$x_i \sim \mathcal{N}_p(0, \Sigma = \Lambda\Lambda^T + \Psi)$$

- Spotlight on sparse estimation of Λ
  $\longrightarrow$ enhance interpretability, detect uninformativeness

$\boxed{\text{What about redundancy?}}$

Factor Analysis with redundant variables

- Idea search for variable clustering structures
  $\longrightarrow$ modify standard FA by allowing some variables to be mapped by means of the same loadings

- Proposed model

$$x_i = Z\Lambda_c u_i + \varepsilon_i \qquad i = 1, \dots, n$$

  - $Z \in \mathbb{R}^{p \times G}$ latent *allocation matrix*, $z_{jg} = 1$ if j-th variable belongs to g-th cluster, G the number of variable clusters
  - $\Lambda_c \in \mathbb{R}^{G \times K}$ cluster specific loadings

- Number of loadings: $(p \times K) \longrightarrow (G \times K)$

Additional thoughts

- In matrix form

$$\begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,p} \end{pmatrix} = \begin{pmatrix} z_{1,1} & \cdots & z_{1,G} \\ \vdots & \ddots & \vdots \\ z_{p,1} & \dots & z_{p,G} \end{pmatrix} \begin{pmatrix} \lambda^c_{1,1} & \cdots & \lambda^c_{1,K} \\ \vdots & \ddots & \vdots \\ \lambda^c_{G,1} & \dots & \lambda^c_{G,K} \end{pmatrix} \begin{pmatrix} u_{i,1} \\ \vdots \\ u_{i,K} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,p} \end{pmatrix}$$

- Uniformativeness can be detected by forcing a row of Λ_c to be equal to a zero vector

- Denoting with $\tilde{\Lambda} = Z\Lambda_c$ and adapting distributional assumptions we obtain

$$(x_i|Z) \sim \mathcal{N}_p(0, \tilde{\Sigma} = \tilde{\Lambda}\tilde{\Lambda}^T + \Psi)$$

Model estimation - Priors & Algorithm

- Bayesian estimation procedure is adopted
  - Standard prior distribution for Λ_c, u_i, Ψ
  - Prior on the allocation matrix $\longrightarrow$ Product Partition Model
    Correspondence between Z and $\mathbf{c} = \{C_1, \ldots, C_G\}$

$$\pi(\mathbf{c}) \propto \prod_{g=1}^{G} \rho(C_g) = \alpha_z^G \prod_{g=1}^{G} (|C_g| - 1)!$$

- Conjugate nature of the priors allows a Gibbs updating scheme
- MH-step to sample Z via modification of the allocation sampler
  - Idea a single move attempts to reallocate a block of
    variables from one group to another
    $\longrightarrow$ Bigger moves, faster exploration
  - Modification aim to propose moves involving *close* clusters

Model selection

- Number of factors K and number of variable clusters G
- Possible solutions:
  - Information criteria (AICM, BICM, BIC-MCMC...)
  - Nonparametric fashion $\longrightarrow$ infinite groups/factors
- Proposal $\longrightarrow$ ad-hoc initialization strategy
  - Idea: multi-step procedure to mimick $\tilde{\Lambda}$ structure via standard FA and model-based clustering strategy
  - Output: $(K_{\text{init}}, G_{\text{init}})$ initial guess to be used as the starting for a local search
  - Rationale: avoid intensive global search since the focus is on covariance reconstruction

Dairy diet MIRS data

- Milk samples from 120 cows collected weekly during summer months (from 2015 to 2017)

- Two different feeding regimens considered:
    - Pasture, cows maintained outdoors on perennial ryegrass and white clover
    - Total mixed ration (TMR), cows mantained indoors, nutrients combined in a single mix

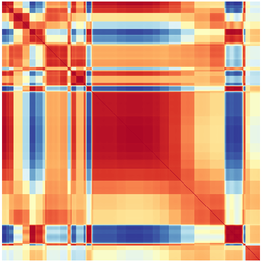
- $n = 4320$ milk samples, $n_P = 2391$ $n_T = 1389$
  $p = 1060$ wavelengths in the mid-infrared region
  $\longrightarrow$ water-absorption spectral regions removed thus $p^* = 533$

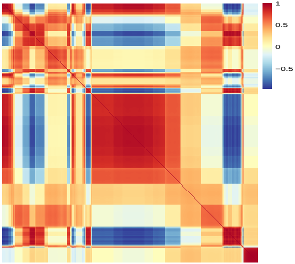
Some results - MIRS data

- Initialization strategy suggests $(K_\mathsf{P} = 4, G_\mathsf{P} = 25)$, $(K_\mathsf{T} = 3, G_\mathsf{T} = 19)$
- Good reconstruction of the sample correlations
- Blocky structure as a byproduct of the clustering mechanism

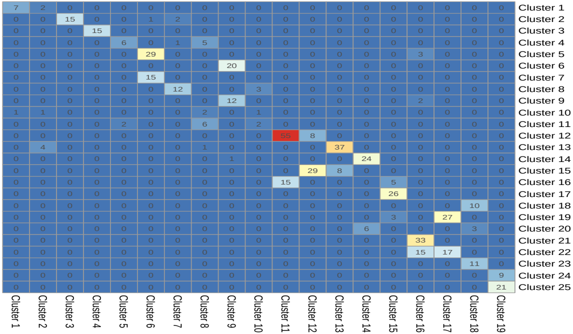
  $\downarrow$

  useful to highlight differences between spectral regions





Some results - MIRS data



- Quite strong agreement between the two partitions (ARI = 0.65) $\longrightarrow$ signal about real and non-spurious clustering structures
- Pronounced redundancy, grouping may be used to build new variables or for cluster-specific predictive analyses
- Interpretability $\longrightarrow$ subject-matter knowledge is needed

Concluding remarks

- The proposed method provides parsimonious summaries of high dimensional data with highly correlated variables
  $\longrightarrow$ not only spectroscopic data
- Richer insights with respect to standard FA thanks to the variable clustering mechanism
- Directions and open questions
  - It might serve as a building block for classification tools, easy to embed in a MFA framework
  - Different choices for the priors?
    $\longrightarrow$ shrinkage priors
    $\longrightarrow$ exploit some *spatial* information when specifying the cohesion function

Some references

Check the paper out on arXiv
https://arxiv.org/pdf/2101.12499.pdf

Other references

- Barry, D. & Hartigan, J.A. (1992) *Product partition models for change point problems*, The Annals of Statistics, 20(1), 260–279.
- Bartholomew, D.J., Knott, M. & Moustaki, I. (2011) *Latent variable models and factor analysis: a unified approach*, John Wiley & Sons.
- Frühwirth-Schnatter, S. & Lopes, H.F. (2010) *Parsimonious Bayesian factor analysis when the number of factor is unknown*, Technical report.
- Nobile, A. & Fearnside, A.T. (2007) *Bayesian finite mixtures with an unknown number of components: the allocation sampler*, Statistics and Computing, 17(2), 147–162.